

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Persaingan bisnis dalam perdagangan dunia melalui ekonomi pasar bebas yang diikuti dengan kemajuan teknologi informasi membawa perusahaan pada tingkat persaingan yang semakin ketat dan semakin terbuka dalam memenuhi tuntutan pelanggan yang juga semakin tinggi. Sehingga memaksa para pelakunya untuk senantiasa memikirkan strategi-strategi dan terobosan bagaimana untuk tetap bertahan dan mengembangkan pangsa pasar bisnis mereka.

212 *Mart* Jalan Rambutan Kota Pekanbaru merupakan salah satu perusahaan yang bergerak dibidang retail. Memenuhi kebutuhan konsumen setiap harinya dan mengambil keputusan yang tepat dalam menentukan strategi penjualan merupakan suatu keharusan. Untuk itu, dibutuhkan sumber informasi yang cukup banyak untuk dapat dianalisis lebih lanjut. Pada bisnis retail, salah satu cara untuk mengetahui bagaimana kondisi pasar (pelanggan) saat ini adalah melakukan pengamatan data transaksi penjualan. Data transaksi penjualan ini disimpan dalam basis data *server* yang kemudian diolah sehingga menghasilkan sebuah informasi berupa laporan penjualan dan laporan laba rugi. Namun, pengolahan data lebih lanjut sangat dibutuhkan agar dapat menghasilkan informasi baru (Erwin, 2009).

Pengolahan data yang lebih lanjut, dapat dilakukan dengan proses ekstraksi dan identifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari data dengan jumlah yang besar (*database*) dengan menggunakan teknik-teknik tertentu seperti teknik *statistic*, matematika, kecerdasan buatan, dan *machine learning* yang biasa disebut dengan *Data mining* (Mustakim, 2012). Banyak metode yang terdapat pada *data mining*, salah satunya yaitu *Association Rule*.

Aturan asosiasi (*Association Rule*) adalah salah satu metode *data mining* yang digunakan untuk menemukan pengetahuan dari sejumlah data yang besar yang terdapat dalam basis data (Defit, 2012). *Association Rule* bertujuan untuk menemukan *frequent itemset* dalam *database* dengan menggunakan *minimum support* dan membuat sebuah aturan asosiasi dari *frequent itemset* dengan *confidence* tertentu. Dengan ketentuan setiap terjadinya kejadian A, maka terjadi kejadian B (Abaya, 2012). *Association Rule* biasa digunakan untuk menganalisis keranjang pasar (*Market Basket Analysis*) (Christidis, Apostolou, dan Mentzas, 2010).

Analisis keranjang pasar (*Market Basket Analysis*) dapat membantu para pelaku usaha retail dalam memberikan dukungan keputusan ilmiah dengan melakukan *Association Rule Mining* antara barang-barang yang dibeli secara

bersamaan oleh pelanggan (Wen-xiu, Heng-nian, dan Mei-li, 2010). Hal ini dapat membantu memberikan rekomendasi dan promosi produk, sehingga menghasilkan strategi pemasaran yang lebih tepat sasaran dan menarik lebih banyak pelanggan karena barang yang dipromosikan merupakan barang yang dibutuhkan oleh pelanggan. Penentuan pola pembelian yang kurang akurat dapat mengakibatkan kebijakan rekomendasi dan promosi produk menjadi tidak tepat sasaran (Triyanto, Suhartono, dan Himawan, 2014).

Pada perusahaan 212 *Mart* Jalan Rambutan Kota Pekanbaru, penentuan promosi produk didapatkan dari analisis laporan data transaksi penjualan, yaitu berdasarkan pada produk yang paling laku. Data transaksi penjualan ini, dievaluasi 1 kali dalam 3 bulan. Sehingga, hasil yang didapatkan tidak efisien. Karena, untuk produk yang paling laku, sering kali promo yang ditawarkan tidak sesuai dengan kondisi. Contohnya, produk yang paling laku ditanggal tanggal tertentu seperti pada awal bulan, ditawarkan di akhir bulan, begitu juga sebaliknya.

Selain itu, penentuan promosi produk juga didapat berdasarkan tanggal kedaluwarsa produk. Produk yang memiliki sisa waktu dalam kurun waktu 3 bulan menjelang kedaluwarsa, akan diadakan promo potongan harga. Sering kali produk-produk yang dipromosikan justru tidak sesuai dengan kebutuhan pelanggan

Sudah banyak algoritma yang diusulkan guna menemukan pola asosiasi (Ykhlef, 2011). Namun, yang paling umum digunakan dalam menentukan *Association Rules* ialah algoritma Apriori. Apriori adalah algoritma aturan asosiasi yang paling terkenal dan digunakan dalam sebagian besar produk komersial (Patil dan Patil, 2013). Algoritma ini berkaitan dengan *set item*, yang kita sebut data transaksi (Wu dan Sakai, 2015). Apriori memiliki kelebihan yaitu penggunaannya yang lebih sederhana, dapat menangani data dengan jumlah yang besar dan hemat memori saat mengolah jumlah data yang besar (Oktoria dan Maharani, 2010). Namun, Apriori juga memiliki kekurangan, dalam pencarian *frequent itemset*, Apriori harus melakukan *scanning database* berulang kali secara keseluruhan tiap kali iterasi (Aggarwal dan Sindhu, n.d.) sehingga apabila data transaksi yang diproses banyak, waktu yang dibutuhkan pun juga banyak (Septiana dan Dharmayanti, 2016). Selain itu, *generate candidate* menjadi hambatan bagi Apriori (Han, Pei, dan Yin, 2001), dibutuhkan *generate candidate* yang besar untuk mendapatkan kombinasi *item* dari *database* (Erwin, 2009).

Beberapa algoritma telah dikembangkan untuk mengatasi kekurangan dari Apriori, seperti algoritma *Frequent Pattern-Growth* (FP-Growth). FP-Growth adalah metode *data mining* pengembangan dari Apriori yang digunakan untuk mencari *frequent itemset* tanpa menggunakan *candidate generation*. Dengan menggu-

nakan konsep pembangunan *tree* dalam pencarian *frequent itemset*, *candidate generation* pun dihilangkan. Hal inilah yang membuat algoritma *FP-Growth* lebih cepat dibandingkan algoritma Apriori (Fitria dkk., 2017). Selain itu, *FP-Growth* hanya melakukan *scanning database* sebanyak dua kali (Kavitha dan Selvi, 2016), dengan demikian *FP-Growth* dapat memperbaiki dua kelemahan dari algoritma Apriori.

Beberapa penelitian yang dilakukan terkait dengan *Association Rules*, seperti Al-Maolegi dan Arkok (2014) yang berjudul *An Improved Apriori Algorithm For Association Rules*. Bertujuan untuk mengurangi waktu yang terbuang begitu banyak dalam proses *scanning database* dalam mencari *itemset* yang sering muncul pada Apriori.

Pada penelitian ini, solusi yang diberikan adalah dengan mengurangi jumlah transaksi yang akan di *scanning*. Setiap kali k dari k -*itemset* meningkat, kesenjangan waktu yang dibutuhkan antara Apriori yang ditingkatkan dan Apriori yang asli sangat meningkat, dan ketika nilai *support minimum* meningkat, kesenjangan waktu yang dibutuhkan antara Apriori yang ditingkatkan dan Apriori asli menurun. Apriori yang telah ditingkatkan mengurangi konsumsi waktu sebesar 67,38%.

Persamaan dengan penelitian yang akan dilakukan ialah penerapan teknologi *data mining* dengan metode *Association Rules*. Namun, perbedaannya adalah tujuan penggunaan metode atau teknik analisis datanya, pada penelitian sebelumnya bertujuan untuk meningkatkan performa Apriori dengan mengurangi waktu *scanning*. Sedangkan pada penelitian yang akan dibuat bertujuan untuk memberikan rekomendasi penawaran produk bagi pelanggan.

Penelitian lainnya yang berjudul *Data mining in Market Basket Transaction: An Association Rule Mining Approach* dilakukan oleh Abdulsalam (2014) dan penelitian yang dilakukan oleh Ramdhan, Achmad, dan Mutaqin (2016) dengan judul *Market Basket Analysis Menggunakan Algoritme Apriori* dan penelitian yang dilakukan oleh Karthiyayini dan Balasubramanian (2016) yang berjudul *Affinity Analysis and Association Rule Mining using Apriori Algorithm in Market Basket Analysis* yang bertujuan untuk menentukan produk mana yang sering dibeli oleh pelanggan secara bersamaan dari data transaksi penjualan sehingga membantu para pelaku retail dalam mengambil keputusan.

Persamaan dengan penelitian yang akan dilakukan ialah penerapan teknologi *data mining* pada data transaksi penjualan guna menemukan sebuah informasi atau pengetahuan baru yang berguna untuk membantu pelaku usaha retail dalam mengambil keputusan sehingga dapat meningkatkan penjualan perusahaan. Namun, perbedaannya adalah penggunaan algoritma dalam analisis datanya, pada penelitian sebelumnya menggunakan Apriori. Sedangkan pada penelitian yang akan dibuat

menggunakan *FP-Growth*.

Penelitian yang lainnya yang berjudul *Market Basket Analysis Using Apriori and FP-Growth for Analysis Consumer Expenditure Patterns at Berkah Mart* in Pekanbaru Riau dilakukan oleh Mustakim dkk. (2018) yang bertujuan untuk menentukan tata letak dan perencanaan ketersediaan / stok barang.

Pada penelitian ini, solusi yang diberikan adalah dengan melakukan analisis keranjang kasar dengan menerapkan teknik asosiasi yaitu dengan algoritma Apriori dan *FP-Growth*, yang mana nantinya kedua metode tersebut akan dibandingkan. Dari kedua algoritma tersebut, terdapat dua hal yang ingin diketahui dari analisis *dataset* penjualan pada *Berkah Mart*, yaitu proses *rule* yang akan dihasilkan dari masing-masing algoritma dan waktu yang dicapai masing-masing algoritma dalam membentuk *rule*.

Dalam menganalisis keranjang pasar menggunakan *Association Rules*, ditemukan sejumlah masalah teknis umum yang berhubungan dengan teknik rekomendasi. *Itemset* besar cenderung diabaikan oleh aturan asosiasi, dan kurang tepatnya rekomendasi *item* dikarenakan tidak tersedianya informasi tentang produk retail (Christidis dkk., 2010), sehingga hasil yang diperoleh menjadi kurang akurat untuk data yang besar. Untuk itu, dilakukan *clustering* terhadap atribut-atribut yang ada guna membentuk kelompok atribut yang sama, setelah itu menentukan pola asosiasi pada masing-masing kelompok yang sudah terbentuk (Plasse, Niang, Saporta, Villeminot, dan Leblond, 2007), sehingga mempermudah dalam proses pencarian rekomendasi produk.

Clustering adalah teknik yang efektif untuk analisis data eksploratori, dan telah diterapkan di berbagai bidang. Secara umum, metode untuk melakukan *clustering* dapat dikategorikan menjadi empat, yaitu: *partitioning*, *hierarchical*, *grid-based and model-based*. Partisi dari data sehingga objek dalam suatu *cluster* lebih dominan mirip dibandingkan objek yang terdapat pada *cluster* lain dihasilkan melalui *clustering* berbasis *partitioning*. Contoh dari metode *partitioning* adalah *K-Means* dan *K-Medoids* (Velmurugan, 2012).

Kelemahan *K-Means* terdapat pada penentuan *centroid* awal pada *cluster*. Hasil *cluster* yang terbentuk tergantung pada inisiasi nilai *centroid* awal *cluster* yang diberikan. Sehingga menyebabkan hasil *cluster* berupa solusi yang bersifat *local optimal* (Rahayu dan Mustakim, 2017). Selain itu, algoritma ini sensitif terhadap *outlier* karena objek dengan nilai yang sangat besar secara substansial dapat mendistorsi distribusi data. Untuk itu, dalam mengambil nilai rata-rata dari objek pada sebuah *cluster* sebagai titik acuan, *medoid* dapat diandalkan, yang mana merupakan objek yang paling terpusat dalam sebuah *cluster*. Menemukan *k cluster*

dalam n objek dengan pertama kali secara *arbitrarily* menemukan wakil dari objek (*medoid*) untuk tiap-tiap *cluster*, merupakan strategi dasar dari algoritma *clustering* K-Medoids (Velmurugan, 2012).

Berdasarkan pembahasan penelitian sebelumnya diatas, pada penelitian ini akan menggunakan algoritma K-Medoids untuk *clustering* pada data penjualan dan menerapkan algoritma FP-Growth untuk pendekatan asosiasi pada setiap *cluster* yang telah dibentuk. Sehingga diharapkan untuk dapat memberikan rekomendasi produk yang lebih akurat kepada pelanggan dikarenakan *dataset* yang akan diasosiasi menjadi lebih kecil.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan diatas, masalah yang ditemukan adalah aturan asosiasi yang dihasilkan menjadi kurang akurat ketika jumlah data yang akan diolah besar, meskipun algoritma FP-Growth sering digunakan untuk menemukan rekomendasi produk. Untuk mengatasi permasalahan tersebut, dilakukanlah proses *clustering* data dengan menggunakan K-Medoids terlebih dahulu sebelum menentukan pola asosiasi. Pertanyaan yang kemudian muncul ialah berapa *rule valid* yang dihasilkan melalui FP-Growth untuk menentukan rekomendasi produk pada *dataset* yang besar apabila diterapkan K-Medoids pada proses *clustering* data?

1.3 Batasan Masalah

Batasan masalah tugas akhir ini adalah:

1. Studi kasus penelitian ini adalah 212 *Mart* Jalan Rambutan Kota Pekanbaru.
2. Data yang digunakan ialah data transaksi penjualan 212 *Mart* Jalan Rambutan Kota Pekanbaru. Total data yang diperoleh yaitu sejak 21 Oktober 2018 sampai dengan 12 April 2019 sebanyak 13827 *record* data sebelum dilakukan *cleaning* dan 13809 *record* data yang digunakan setelah dilakukan *cleaning*.
3. Data transaksi yang digunakan adalah data transaksi penjualan produk produk harian, tidak termasuk didalamnya produk perkakas rumah tangga seperti sapu, pel, Alat tulis kantor seperti pena, pensil, dan lain sebagainya.
4. Pada perhitungan algoritma K-Medoids menggunakan percobaan dengan jumlah *cluster* 3, 4, 5, 6, 7, 8 dan 9 sebagai permodelan percobaan.
5. Acuan yang digunakan dalam pengelompokan *medoid* yaitu berdasarkan atas Jumlah *Item* Yang Dibeli dan Jumlah *Item* Yang Tidak Dibeli.
6. Metode validitas *cluster* yang digunakan adalah *Davies Bouldien Index* (DBI)

7. Pada perhitungan algoritma *FP-Growth* menggunakan percobaan dengan nilai *minimum support* yaitu 5%, 6%, 7%, 8%, 9% dan nilai *minimum confidence* yaitu 50%.
8. *Tools* yang digunakan dalam penelitian Tugas Akhir yaitu RapidMiner dan Excel.

1.4 Tujuan

Tujuan tugas akhir ini adalah menerapkan algoritma K-Medoids untuk *clustering* pada *FP-Growth* dalam menghasilkan *rule* rekomendasi produk pada jumlah *dataset* yang besar.

1.5 Manfaat

Manfaat tugas akhir ini adalah:

1. Membantu pelaku bisnis retail mengambil keputusan dalam membantu menentukan kebijakan, khususnya dalam bidang strategi pemasaran.
2. Mempermudah pelanggan dalam berbelanja, karena rekomendasi produk yang diberikan lebih akurat.
3. Memberikan rekomendasi teknik/cara baru pada pihak 212 *Mart* Jalan Rambutan Kota Pekanbaru dalam penentuan promosi produk.
4. Mencari perbandingan hasil sebelum dilakukannya *cluster* dan sesudah dilakukannya *cluster* dalam penerapan algoritma *FP-Growth*.

1.6 Sistematika Penulisan

Sistematika penulisan laporan Tugas Akhir ini terdiri dari 5 (lima) bab yang disusun untuk memberikan gambaran umum tentang penelitian yang sedang dilakukan, dengan sistematika penulisan sebagai berikut:

BAB 1. PENDAHULUAN

Bab ini berisi penjelasan mengenai: (1) latar Belakang; (2) rumusan masalah; (3) tujuan; (4) manfaat; dan (5) sistematika penulisan.

BAB 2. LANDASAN TEORI

Bab ini berisi teori tentang: (1) rekomendasi; (2) produk; (3) pelanggan; (4) analisis keranjang pasar; (5) *data mining*; (6) *clustering*; (7) algoritma k-medoids; (8) asosiasi (*association rules*); (9) algoritma *fp-growth*; (10) rapidminer; dan (11) penelitian terdahulu.

BAB 3. METODOLOGI PENELITIAN

Bab ini berisi langkah-langkah dan metode yang dilakukan dalam penelitian ini, yaitu: (1) tahap perencanaan; (2) tahap pengumpulan data; (3) tahap *preprocessing*; (4) tahap analisis dan hasil; dan (5) tahap dokumentasi.

BAB 4. ANALISA DAN PERANCANGAN

Bab ini berisi hasil dan pembahasan mengenai hasil: (1) analisis pendahuluan; (2) pengumpulan data; (3) penentuan atribut; (4) *preprocessing* data; (5) pengelompokan data penjualan dengan algoritma k-medoids menggunakan *tools* rapidminer; (6) pencarian *association rules* data penjualan yang telah di *cluster* dengan algoritma *fp-growth* menggunakan *tools* rapidminer; (7) pencarian *association rules* data penjualan tanpa *cluster* dengan algoritma *fp-growth* menggunakan *tools* rapidminer; dan (8) perbandingan hasil *association rules* setelah di *cluster* dengan sebelum di *cluster*.

BAB 5. IMPLEMENTASI DAN PENGUJIAN

Bab ini berisikan tentang kesimpulan dan saran dari hasil Tugas Akhir.